

複数の特徴量間の関連性に着目した Joint-HOG による物体検出

藤吉 弘亘, 山内 悠嗣, 三井 相和, 池村 翔 (中部大学), 山下 隆義 (中部大学, オムロン株式会社)

Object Detection by Joint-HOG Based on Relations between Multiple Features
Hironobu Fujiyoshi, Yuji Yamauchi, Tomokazu Mitsui, Sho Ikemura (Chubu University)
Takayoshi Yamashita (OMRON Corporation, Chubu University)

Abstract

This paper presents a method for detecting people in a video using a Joint-HOG feature, which represents co-occurrence between multiple HOG features. Capturing the co-occurrence feature of Joint-HOG feature makes it possible to construct the effective human detector. We also describe effectiveness of multi-resolution HOG representation, co-occurrence of spatio-temporal features, and co-occurrence of depth information obtained by a TOF (Time of Flight) camera.

キーワード: 人画像解析, 人検出, HOG, Joint-HOG 特徴量, 共起,
(People Image Analysis, People Detection, Histograms of Oriented Gradients, Joint-HOG feature, Co-occurrence)

1. はじめに

近年, オフィスや家, 公共施設などの空間において, その空間内の人の意図を理解し行動を支援する技術の実現が期待されている. ジョージア工科大学の Aware Home プロジェクト⁽¹⁾では, 生活空間である家にカメラをはじめとする数多くのセンサ群を埋め込み, 24 時間を通して生活空間における人の動きをセンシングする研究に取り組んでいる. また, リビングルームを対象とした Microsoft Research の Easy Living⁽²⁾のように, センシングにより得られた情報を基に, ユーザである人に対して快適な空間をアシストする研究が盛んである. このような活力生活技術は QoLT (Quality of Life Technology) と呼ばれる技術の一環で, QoLT のためのセンシング技術は, 刻々と変化する人の状態を実時間で認識する必要がある. 特に, 人画像解析 (People Image Analysis) として, 動画像からの人の検出, 追跡, 顔の検出, 顔の部位の追跡, モーション理解が不可欠な技術要素となる. 中でも人検出は, 人の形状変化が大きいいため難しい問題とされている. しかし, 近年のコンピュータの高速化に伴い, 画像全体に検出ウィンドウをラスタスキャンし, low-level な特徴量と統計的学習手法の組み合わせを用いることで, 高精度に人を検出する手法が提案されている.

人検出に用いられる局所特徴には, HOG (Histograms of Oriented Gradients) 等のエッジベースの特徴量^{(3) (4) (5)}が多く提案され, 高い検出性能を示すことが報告されている. 本研究の検出対象である歩行者の形状には, 下記に示すように大きく分けて 2 つの特徴がある.

(1) 頭から肩にかけての Ω に似た形状や上半身から下半身にかけての連続的な形状

(2) 頭や肩, 胴, 足などの左右対称的な形状

(1) に対しては, 局所領域内の 4 方向のエッジ特徴を Ad-

aBoost により組み合わせることにより局所領域の面の情報を表現した Shapelet 特徴量⁽⁶⁾がある. (2) に対しては, AdaBoost の弱識別器が複数の特徴量を同時に観測し, 共起を表現する Joint Haar-like 特徴量⁽⁷⁾が提案されている. 両手法は, 複数の low-level な特徴量をブースティングにより組み合わせることにより特徴量間の関連性を捉えることができ, 高精度な検出が可能となる. そのため, 特徴をどのように捉えるかが重要であるが, (1) と (2) の両方の形状特徴を捉えることができる手法は提案されていない.

そこで, 我々は物体形状の対称性や連続性を自動的に捉えることが可能である Joint-HOG 特徴量と 2 段階 AdaBoost による人検出法を提案した. Joint-HOG 特徴量は, 異なる 2 つの領域の HOG 特徴量を AdaBoost により組み合わせることで, 単一の特徴量では表現が困難な対称性や連続的なエッジを 1 段階目の AdaBoost により生成することが可能となる. 生成された Joint-HOG 特徴量を入力とした 2 段階目の AdaBoost によって最終識別器を構築し, 検出ウィンドウ内が人/人以外に判別する. 本稿では, 我々が取り組んできた Joint-HOG 特徴量による人検出法と特徴量間の共起による効果について述べる.

2. 局所特徴量とブースティング

従来の物体検出には, 入力画像と背景画像の差分を計算する背景差分ベースの手法が多く用いられてきた. このような動体検出をベースとした動画像理解のアプローチは, 移動体同士が画像上で重なった場合にセグメンテーションに失敗するため, その後の処理である物体識別が不可能となる問題がある.

これに対して, Viola と Jones は Haar-like 特徴量と呼ばれる局所特徴量と統計的学習の組み合わせによる高速かつ高精度な顔検出法⁽⁸⁾を提案した. この手法は, 入力画像に

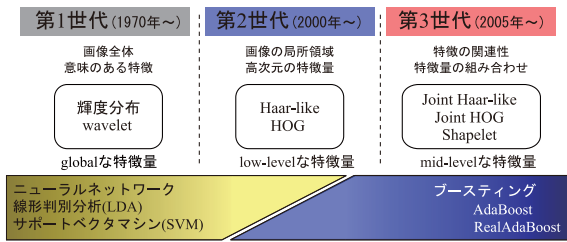


図 1 学習法と特徴量の変遷

Fig. 1. Learning method transition and features.

対して検出ウィンドウをラスタスキャンし、AdaBoost を用いて検出ウィンドウ内の画像が顔/顔以外として判別するため、前処理として動体検出を必要としない。顔検出の分野では、ラスタスキャン方式による手法が多く提案され実用化された。近年では、検出対象が顔から形状変化が大きいため検出がより難しいとされている人へと変わりつつある。本章では、近年の人検出のアプローチである局所特徴量と統計的学習の変遷と、HOG 特徴量を用いた人検出法について述べる。

〈2・1〉 識別器と特徴量の変遷

物体検出に用いられる特徴量は図 1 に示すように、第 1 世代の global な特徴量、第 2 世代の low-level な特徴量、第 3 世代の mid-level な特徴量に分けることができる。以下に各世代毎の特徴量について述べる。

第 1 世代 物体検出に用いられる第 1 世代の特徴量は、輝度分布や wavelet⁽⁹⁾ など画像全体から得られるものや、形状の複雑度のような意味のある特徴量を研究者の経験に基づき決定していた。第 1 世代の識別器としては、ニューラルネットワークが用いられており、次元数が大きな特徴量では、ニューラルネットワークのサイズが大きくなり、構築が困難となる。そのため、入力ユニットの数を小さくする必要があり、低次元数の特徴量が用いられていた。

第 2 世代 2000 年以降に提案された第 2 世代では、Haar-like 特徴量⁽⁸⁾ や HOG 特徴量などの low-level な局所特徴量から、統計的学習法であるブースティングを用いて特徴選択を行うため、高次元 (数千～数十万パターン) の特徴量を扱うことが可能となった。これにより、第 1 世代では研究者が特徴量を決定していたが、第 2 世代では研究者の事前知識なく、多数の特徴量候補の中から有効な特徴量を自動的に選択することができる。

第 3 世代 第 3 世代 (2005 年以降) では、第 2 世代の low-level 特徴量をブースティングの特徴選択を利用して組み合わせ、識別に有効な関連性や共起性を表現した Joint Haar-like 特徴量⁽¹⁰⁾ や Joint-HOG 特徴量⁽¹¹⁾、Shapelet 特徴量⁽⁶⁾ などが登場した。これらは、low-level 特徴量である第 2 世代の特徴量をベースとし、1 つ 1 つでは意味のない low-level な特徴量に対して、関連性を考慮することで何らかの意味をもつ mid-level な特徴量に拡張している。

〈2・2〉 HOG 特徴と人検出

人や車等の物体検出に

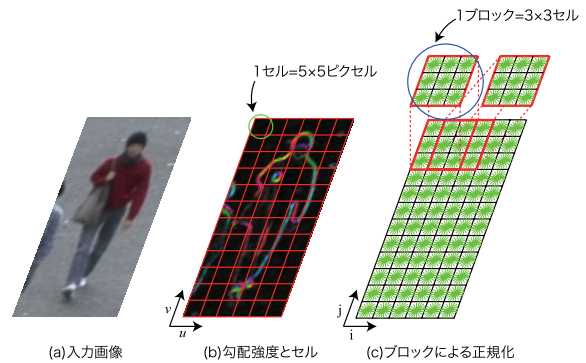


図 2 HOG で用いる領域の構造

Fig. 2. Structure of region with HOG.

は、第 2 世代の局所特徴量である Histograms of Oriented Gradients(HOG)⁽³⁾ が多く用いられている。HOG 特徴量は、局所領域における輝度の勾配方向をヒストグラム化した特徴量であり、ある一定領域に対する特徴量の記述を行う。そのため、大まかな物体形状を表現することが可能であり、人検出⁽³⁾⁽¹²⁾ や車検出⁽¹³⁾ 等の物体検出に有効であることが報告されている。

〈2・2・1〉 HOG 特徴量の算出アルゴリズム

HOG 特徴量の算出は、画像から輝度勾配を算出し、次に、勾配強度と勾配方向から輝度の勾配方向ヒストグラムを作成後に正規化を行う。以下に HOG 特徴量算出アルゴリズムについて述べる。

まず、各ピクセルの輝度から勾配強度 m と勾配方向 θ を次式より算出する。

$$m(u, v) = \sqrt{f_u(u, v)^2 + f_v(u, v)^2} \dots\dots\dots (1)$$

$$\theta(u, v) = \tan^{-1} \frac{f_v(u, v)}{f_u(u, v)} \dots\dots\dots (2)$$

$$\begin{cases} f_u(u, v) = I(u+1, v) - I(u-1, v) \\ f_v(u, v) = I(u, v+1) - I(u, v-1) \end{cases} \dots\dots\dots (3)$$

次に、図 2 に示すように、5 × 5 ピクセルをセルとした領域において輝度の勾配方向ヒストグラムを作成する。輝度の勾配方向ヒストグラムは、0° - 180° を 20° ずつに分割するため、9 方向の勾配方向ヒストグラムとなる。

最後に、各セルで作成した輝度の勾配方向ヒストグラムを 3 × 3 セルを 1 ブロックとして正規化を行う。 i 行 j 列のセル (i, j) の特徴量 (9 次元) を $F_{ij} = [f_1, f_2, \dots, f_9]$ とすると、 k 番目のブロックの特徴量 (81 次元) は $V_k = [F_{i \ j}, F_{i+1 \ j}, F_{i+2 \ j}, F_{i \ j+1}, F_{i+1 \ j+1}, F_{i+2 \ j+1}, F_{i \ j+2}, F_{i+1 \ j+2}, F_{i+2 \ j+2}]$ と表すことができる。正規化後の特徴量を v としたとき、次式より正規化する。

$$v = \frac{f}{\sqrt{\|V_k\|_2^2 + \epsilon^2}} \quad (\epsilon = 1) \dots\dots\dots (4)$$

正規化は、図 2(c) のようにブロックを 1 セルずつ移動させることによって正規化を行う。そのため、特徴量 f は異なるブロックの領域によって何度も正規化される。入力画像

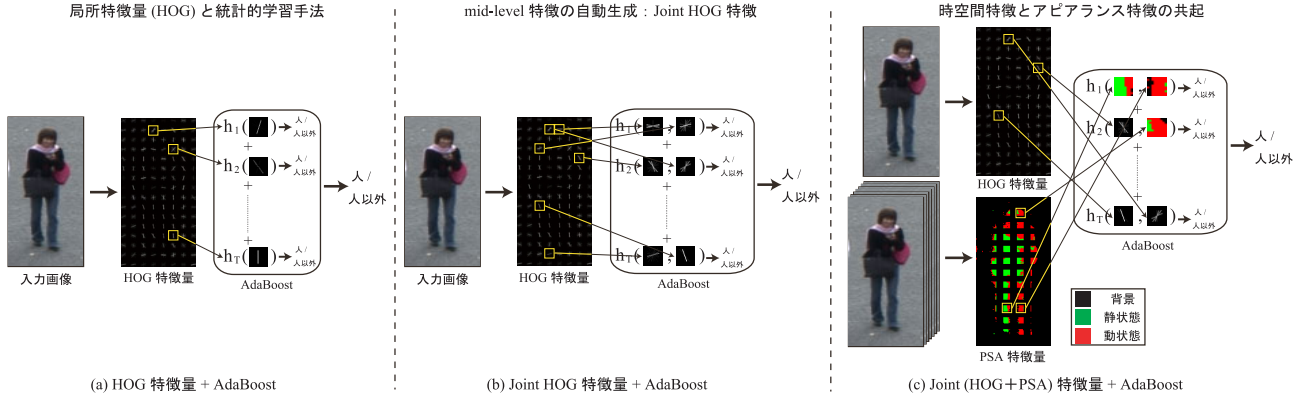


図 3 人検出における特徴量の捉え方

Fig. 3. Feature representation for detecting people.

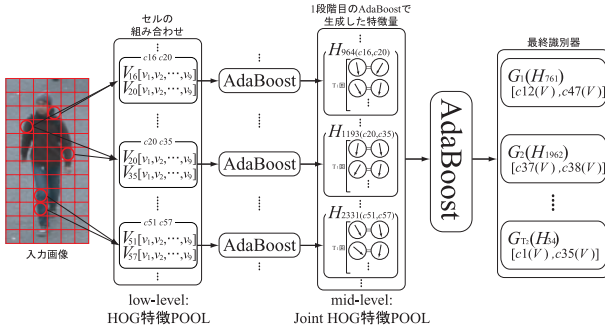


図 4 Joint-HOG 特徴量による 2 段階 AdaBoost

Fig. 4. Two-stage AdaBoost with Joint-HOG.

を 30×60 ピクセルとした場合、各ブロックごとに正規化された HOG 特徴量は、3,240 個となる。

〈2・2・2〉 ブースティングによる人検出 人は動きとともに形状が変化する非剛体な物体であるため、人の検出は顔検出と比較して難しい問題である。また、オクルージョンの発生や衣服の違い、照明や影の影響も検出を困難とする要因である。このような問題に対して、Dalal 等により HOG 特徴量と統計的学習手法を組み合わせた人検出法⁽³⁾が提案された。HOG 特徴量は、照明の変動による影響が少なく、局所的な幾何学的変化に頑健であるため高精度な人検出を可能とした。

図 3(a) に統計的学習手法である AdaBoost による HOG 特徴量の捉え方を示す。学習後の AdaBoost の弱識別器により 1 個の HOG 特徴量が自動で選択され、最終的に多数ある弱識別器の重み付き多数決により人と人以外に判別することが可能となる。このような識別器を画像左上から検出ウィンドウをスケール変化させ複数回ラスタスキャンを行い、各検出ウィンドウに対して人であるかを識別する。人として識別されたウィンドウは、最後に Mean Shift クラスタリング⁽¹⁴⁾ 等によるウィンドウの統合処理を行う。

3. Joint-HOG 特徴量と共起

人には、形状の左右対称性や連続したエッジがあり、こ

れらの特徴を捉えることで検出精度を向上させることができると考えられる。我々は、人独特の形状を捉えるために、複数の HOG 特徴量を組み合わせた Joint-HOG 特徴量と、2 段階に構築した AdaBoost による学習法⁽¹¹⁾を提案した。以下に、Joint-HOG 特徴量について述べる。

〈3・1〉 Joint-HOG 特徴 Joint-HOG 特徴量は、二つの局所領域内の複数の low-level な特徴量である HOG 特徴量を、AdaBoost により組み合わせた mid-level な特徴量である。Joint-HOG 特徴量を用いた人検出は、Joint-HOG 特徴量を入力とした 2 回目の AdaBoost により識別に有効な Joint-HOG 特徴量を自動的に選択して、最終識別器を構築する(図 4)。図 3(b) に、AdaBoost による Joint-HOG 特徴量の捉え方を示す。HOG 特徴量では、1 個の弱識別器が 1 個の HOG 特徴量を用いて識別したのに対し、Joint-HOG 特徴量では、1 個の弱識別器が位置の異なる 2 つの領域内に含まれる複数の HOG 特徴量を用いて識別を行う。これにより、従来の単一の HOG 特徴量のみでは捉えることができない物体形状の対称性や連続的なエッジを自動的に捉えることができ、高精度な人検出が可能となる。図 5(a) に人の平均勾配画像、図 5(b), (c) に AdaBoost により選択された HOG 特徴量を可視化した結果を示す。HOG 特徴量の勾配方向を 9 方向で表現しており、輝度が高いほど AdaBoost における弱識別器の重みが高いことを表す。1 段階目で選択された HOG 特徴量(図 5(b))は全ての局所領域において選択されているが、2 段階目で選択された HOG 特徴量(図 5(c))では、人の輪郭に沿った HOG 特徴量が自動的に選択され、高い重みを持つことがわかる。

〈3・2〉 多重解像度間の共起 人検出に有効な形状を Joint-HOG 特徴量により捉えることで検出精度を向上させることができた。しかし、人の頭や肩などの形状は、特徴を最も最適に表現することができる解像度が同一であるとは限らない。そこで、入力画像の大きさを変えながら HOG 特徴量を抽出することで、異なる解像度間の共起を表現する。解像度を変化させることで、人の頭や肩などの部位毎で最適な解像度の HOG 特徴量を選択することができる。図

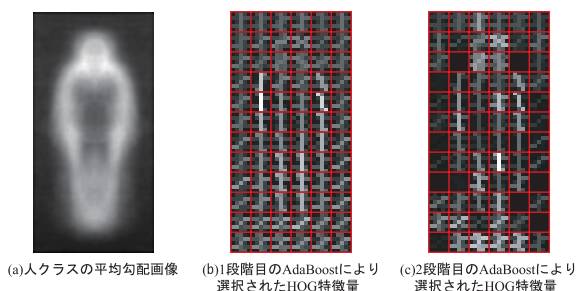


図 5 選択された HOG 特徴量の可視化
Fig. 5. Visualization of selected HOG features.

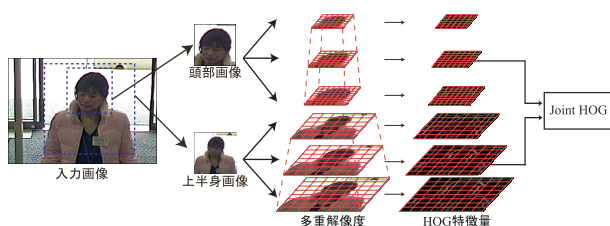


図 6 多重解像度 HOG 特徴量の共起
Fig. 6. Co-occurrence of HOG features of multiple resolution images.

6 に、多重解像度間の共起表現の流れを示す。まず、入力画像から頭部と上半身にわけ、それぞれダウンサンプリングすることで多重解像度画像を作成し、HOG 特徴量を抽出する。そして、各セルより得られた HOG 特徴量を用いて Joint-HOG 特徴量を計算する。

図 7(a), (b), (c) は AdaBoost の各ラウンドにおいて選択された HOG 特徴量の可視化画像、図 7(d) は各学習ラウンドにおいて選択された HOG 特徴量の累積を可視化した画像である。特徴選択の傾向が出やすい学習の初めでは、頭部は高解像度の HOG 特徴量が選択され、上半身は低解像度の HOG 特徴量が選択されていることがわかる。これは、学習サンプルの平均勾配画像からもわかるように、頭部の勾配はばらつきが少ないため高解像度の HOG 特徴量が選択され、上半身の勾配が分散していると低解像度の HOG 特徴量が選択されるためである。低解像度画像からの HOG 特徴量は、広い範囲でヒストグラムをまとめることになるため、このばらつきを吸収することができたと考えられる。また、2 回目の弱識別器では左頭部と右の肩のライン、3 回目の弱識別器では右頭部と左の肩のラインを捉えている。これらのことから、多重解像度において対称性を捉えることにより、部分的なオクルージョンに影響を受けない特徴量が選択されている。

〈3・3〉 時空間特徴との共起 Joint-HOG 特徴量のフレームワークでは、人のアピランスを表す HOG 特徴量 (図 8(b)) に、他の特徴を追加することが可能である。我々は、従来より動体検出に用いられてきた、時空間特徴に基づく特徴量として図 8(c) に示すピクセル状態分析 (PSA) の結果を加えることにより、さらに高精度な人検出を実現

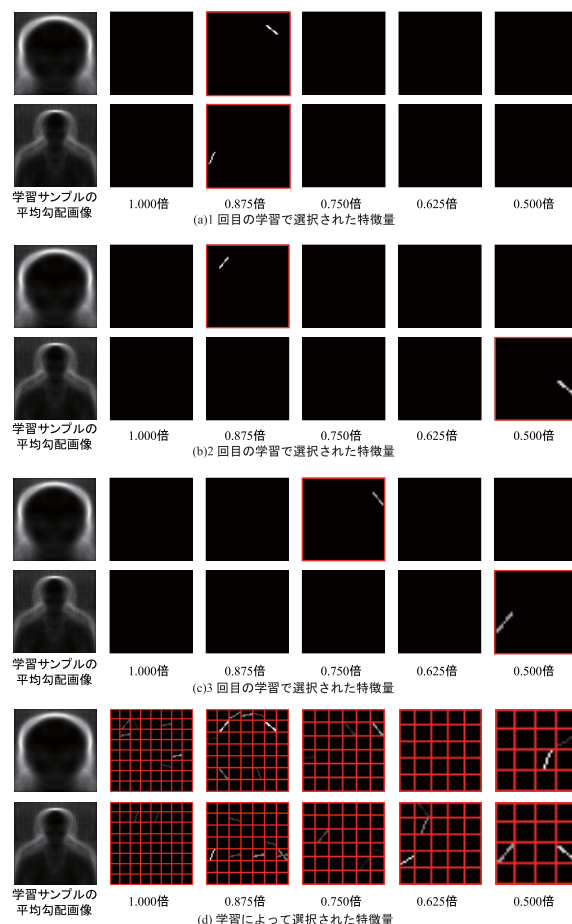


図 7 選択された HOG 特徴量の可視化
Fig. 7. Visualization of selected HOG features.

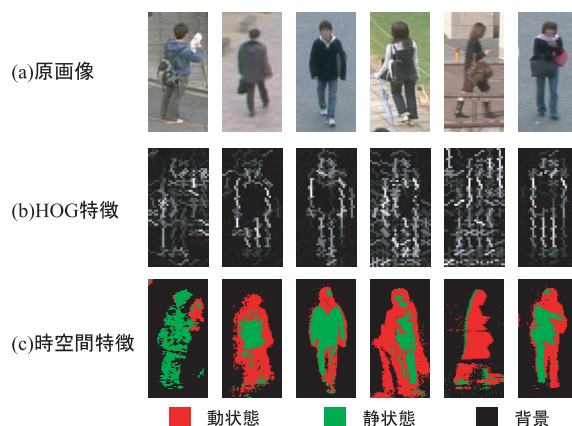


図 8 HOG 特徴とピクセル状態分析の例
Fig. 8. Example of HOG features and pixel state analysis.

した⁽¹²⁾。ピクセル状態分析とは、ピクセル状態の時間変化をモデル化し、各ピクセルを背景と動状態、静状態に判別する手法である。この時空間特徴とアピランス特徴を図 3(c) に示すように同時に捉える手法を示す。時空間特徴を

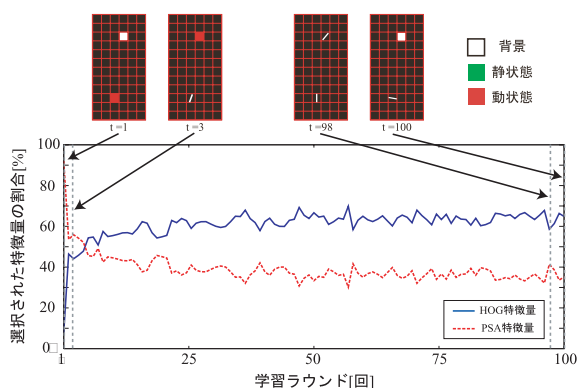


図 9 選択された特徴量の割合
Fig.9. Ratio of selected features.

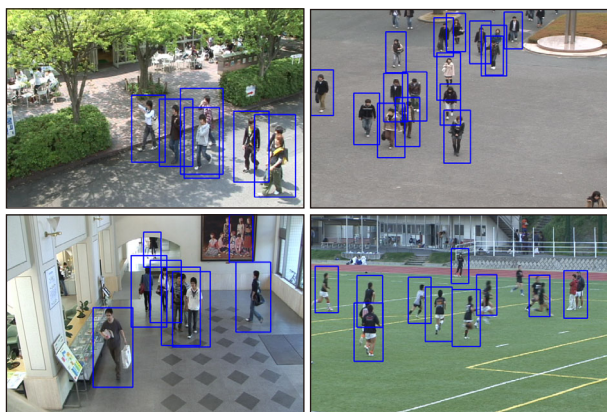


図 10 人検出例
Fig.10. Example of people detection.

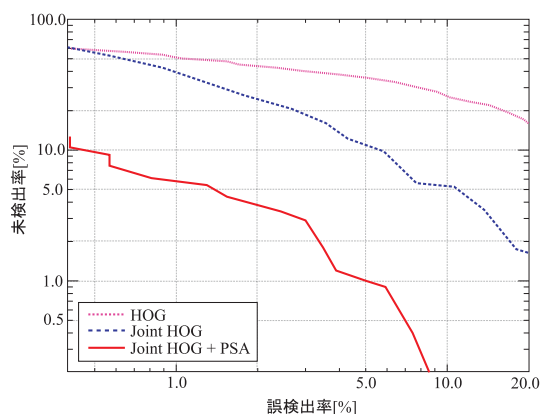


図 11 DET カーブ
Fig.11. DET curve.

加えることにより、人のアピランスと動きの情報を捉えることが可能となる。これにより、アピランスの情報のみでは誤検出する人に似た物体に対して誤検出を抑制することができる。

AdaBoost により選択された特徴量に着目するために、図 9 に各学習ラウンドにおける HOG 特徴量と PSA 特徴量の

選択された割合と、その際に選択された特徴量の可視化の例を示す。初期ラウンドでは PSA 特徴量が多く選択され、学習ラウンド数が進むにつれて HOG 特徴量が選択される割合が多い。これは、まず物体の動きを表すことが可能な PSA 特徴量により、大まかに人と人以外を判別し、この後アピランスの情報をもつ HOG 特徴量を用いて、より細かな識別境界を形成していると考えられる。

図 10 に本手法による人検出の例を示す。人の大きさの変化や人同士の画像上での重なりによる部分的な隠れに対しても高精度な人検出が可能である。図 11 に実験結果の DET カーブを示す。従来人検出に用いられている HOG 特徴量のみよりも、Joint-HOG 特徴量の方が識別性能が高いことがわかる。また、人のアピランス特徴と時空間特徴量を同時に捉えることで、さらに高精度な人検出ができています。従来の HOG 特徴量に比べて Joint-HOG+PSA 特徴量は、誤検出率 5.0%において検出率を 99%まで向上させることができた。

〈3・4〉 デプス情報との共起 可視光カメラにより取得した画像から人の検出を行う場合、オクルージョンや背景のテクスチャの複雑さによって、人検出に有効なアピランス情報を取得することが困難となる場合がある。そこで我々は、カメラから物体までのデプス情報を取得できる Time of Flight カメラ (TOF カメラ) を用いた人検出を提案している。TOF カメラとは、カメラの周囲に付いた複数の LED より照射される赤外光が物体に反射し、カメラで観測されるまでの時間を計測することにより、物体までのデプスを取得するカメラである。この TOF カメラから得られるデプス情報と、アピランス特徴を同時に捉えることにより高精度な人の検出を行う。デプス情報を加えることにより、AdaBoost の弱識別器は物体と背景の距離関係と物体のアピランスを捉えることが可能となり、オクルージョンや背景の複雑さの影響を抑制することができる。

デプス情報から得られる特徴量は、図 12 に示すように、距離画像をセル分割し、注目セルの距離ヒストグラムと 8 近傍のセルの距離ヒストグラムとの Bhattacharyya 距離による類似度を算出し、各セルとの距離を特徴量とする。この特徴量は注目セルに対する近傍セルの相対的な距離関係を表現した特徴量となっている。図 13 に、アピランス情報である HOG 特徴量と距離ヒストグラムから得られる特徴量の共起による人検出の流れを示す。これにより図 14 に示すように、人の重なりがある場合でも高精度な検出が可能となる。

4. おわりに

本稿では、物体検出として局所特徴である HOG 特徴量と、図 15 に示すように、複数の HOG 特徴量を組み合わせた Joint-HOG 特徴について述べた。これらの特徴を統計的学習手法である AdaBoost を用いて学習することにより、精度のよい人検出が可能であることを示した。また、時空間特徴である PSA 特徴量や距離ヒストグラム特徴量と共起させることにより、さらに高精度な人検出を実現するこ

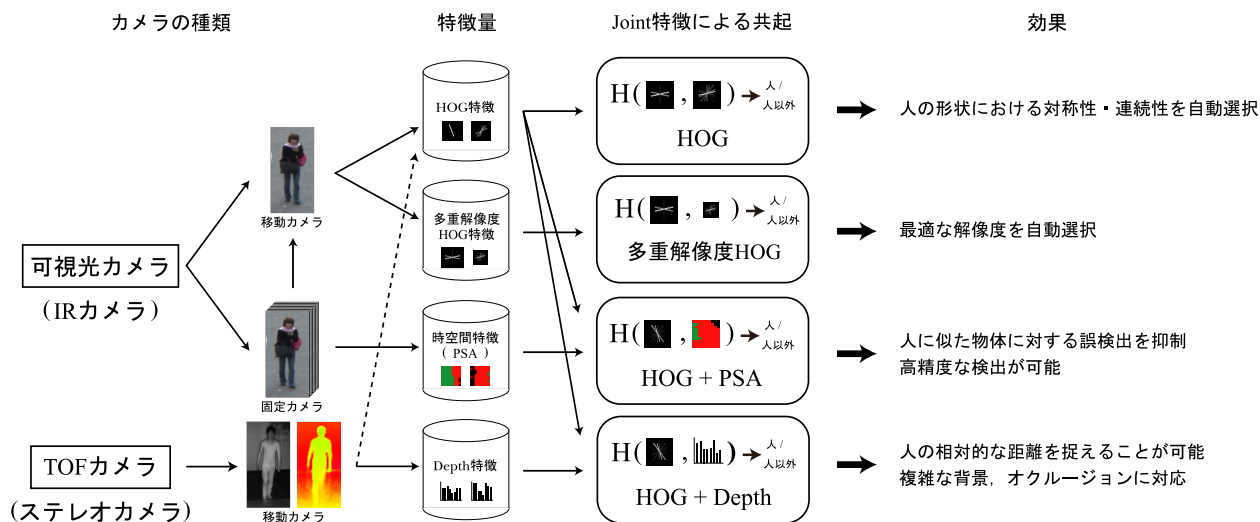


図 15 Joint-HOG による共起特徴表現

Fig. 15. Co-occurrence feature representation by Joint-HOG.

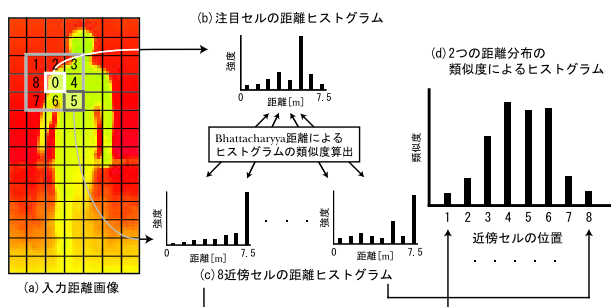


図 12 距離ヒストグラムから得られる特徴量

Fig. 12. Obtained feature from depth histogram.

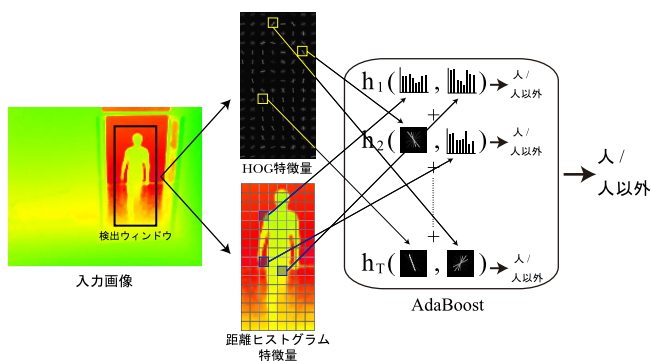


図 13 HOG 特徴量と距離ヒストグラム特徴量の共起

Fig. 13. Co-occurrence of depth and appearance information.

とができた。今後は、ブースティングにおける学習の効率化について検討する予定である。

参考文献

- (1) Aware House,
<http://www.gatech.edu/innovations/futurehome/>
- (2) Easy Living,
http://research.microsoft.com/easy_living/

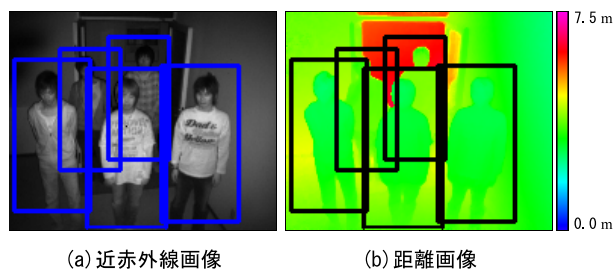


図 14 TOF カメラによる人検出例

Fig. 14. Example of people detection using TOF camera.

- (3) N. Dalal and B. Triggs: "Histograms of Oriented Gradients for Human Detection", CVPR, pp. 886-893, (2005).
- (4) K. Levi and Y. Weiss: "Learning Object Detection from a Small Number of Examples: the Importance of Good Features.", CVPR, vol. 2, pp. 53-60, (2004).
- (5) B. Wu and R. Nevatia: "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", ICCV, vol. 1, pp. 90-97, (2005).
- (6) P. Sabzmeydani and G. Mori: "Detecting Pedestrians by Learning Shapelet Features", CVPR, pp. 1-8, (2007).
- (7) T. Mita, T. Kaneko and O. Hori: "Joint Haar-like Features for Face Detection", ICCV, pp. 1619-1626, (2005).
- (8) P. Viola and M. Jones: "Robust Real-Time Face Detection", IJCV, 57(2), pp. 137-154, (2004).
- (9) H. Schneiderman and T. Kanade: "A statistical method for 3d object detection applied to faces and cars", CVPR, pp. 746-751, (2000).
- (10) T. Mita, T. Kaneko, B. Stenger and O. Hori: "Discriminative Feature Co-occurrence Selection for Object Detection", PAMI, vol. 30, no. 7, pp. 1257-1269, (2008).
- (11) T. Mitsui, Y. Yamauchi and H. Fujiyoshi: "Human Detection by Two Stages AdaBoost with Joint HOG", SSII, IN1-06, (2008).
- (12) 山内悠嗣, 藤吉弘亘, Bon-Woo Hwang, 金出武雄: 「アピランスと時空間特徴の共起に基づく人検出」, 画像の認識・理解シンポジウム (MIRU2007), (2007).
- (13) F. Han, Y. Shan, R. Cekaner, H. S. Sawhney and R. Kumar: "A two-stage approach to people and vehicle detection with HOG-based SVM", Workshop on PMIS, pp. 133-140, (2006).
- (14) D. Comaniciu and P. Meer, "Mean shift analysis and applications", IJCV, pp. 1197-1203, 1999.